EXHIBIT A

Molecular Immunology, Vol. 23, No. 9, pp. 951–960, 1986
Printed in Great Britain.

0161-5890/86 $3.00 + 0.00
Pergamon Journals Ltd

ANTIBODY Fab ASSEMBLY: THE INTERFACE RESIDUES BETWEEN CH1 AND CL

EDUARDO A. PADLAN, GERSON H. COHEN and DAVID R. DAVIES
Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.

(*First received* 24 *January* 1986; *accepted in revised form* 25 *March* 1986)

Abstract—The effective assembly of an antibody molecule requires the proper association of the light and heavy chains, namely the tight, canonical association of VH with VL, and of CH1 with CL. In this paper the interaction of CH1 is examined by looking at the degree of conservation of residues in the interface between CH1 and CL, where CH1 can belong to any of the heavy chain classes, and CL can be either lambda or kappa. The three-dimensional structures of four antibody Fabs have been examined to see which are the significant interacting residues and to see whether they also correspond to the conserved residues in the different classes. It was found that there are a few hydrophobic residues buried in the interface which make numerous contacts with residues of the other chain and which remain invariant, or else are highly conserved. Around the periphery of the interface there are numerous interacting residues that have appreciable variability. Within the interface there is a cavity, the function of which may be to permit some changes in the central interface residues while still preserving the same relative orientation of CH1 and CL.

INTRODUCTION

In an antibody-producing cell large amounts of antibody light and heavy chains are synthesized and are assembled into the complete four-chain antibody molecule. The work of Dorrington and his colleagues (Klein *et al.*, 1979) has demonstrated that the proper formation of the Fab requires the interaction of the two constant domains, CH1 and CL. However, each CL, whether lambda or kappa, must be capable of combining effectively with each CH1, whether the CH1 is alpha, gamma, delta, epsilon or mu. The simplest way of achieving this result would be to have the sequence of the domains, or at least the interface residues of the domains, remain invariant through the different isotypes, but a quick examination of the observed sequences shows that this is not the case.

In order to identify the interacting residues in the interface between CL and CH1, the crystal structures of four Fabs have been examined. The antibodies in these crystals consisted of two human IgG lambda (KOL and NEW) and two mouse IgA kappa (McPC603 and J539) molecules. The results of this examination were then extended to the other classes by aligning the sequences with those of these four Fabs. The degree of conservation of these interface residues in the different classes was then determined. In the following text the results of this investigation are presented and some unexpected observations that were made are described.

MATERIALS AND METHODS

Atomic co-ordinates

The atomic co-ordinates for McPC603 were from a structure refined at 2.7 Å resolution (Satow *et al.*, 1986); those for J539 were from a refinement at 2.6 Å resolution (Suh *et al.*, in preparation). The atomic co-ordinates for NEW (Saul *et al.*, 1978) and KOL (Marquart *et al.*, 1980) were obtained from the Protein Data Bank (Bernstein *et al.*, 1977).

Protein sequences

The amino acid sequences were obtained from Kabat *et al.* (1983). The human CH1 sequences (Table 1) were those for EU (Sequence No. 1), TRO (No. 34), WAH (No. 42), IgE'CL (No. 31) and GAL (No. 38) for human IgG, IgA, IgD, IgE and IgM, respectively. The corresponding mouse sequences that we use here were those obtained from translation of the nucleotide sequence of cloned genomic DNA, specifically sequence Nos 45, 60, 73, 70 and 65 for IgG, IgA, IgD, IgE and IgM, respectively.

The human CL sequences (Table 2) were those for TI (Sequence No. 1) and for NEWM (No. 16) for human kappa and lambda chains, respectively. The corresponding mouse sequences were those for MOPC21 (No. 23) and PLAI-13'CL (No. 31). The numbering scheme of Kabat *et al.* (1983) is used throughout this paper.

Structural comparisons

The CL:CH1 pairs of domains were structurally aligned by a least-squares superposition of the various pairs with those of McPC603. This was accomplished using program ALIGN (G. H. Cohen, unpublished) which is described elsewhere (Satow *et al.*, 1986). Only main chain atoms were used here in the structural alignments.

Sequence alignment

The isolated CL and CH1 domains of McPC603 and KOL were superposed using program ALIGN

Table 1. Amino acid sequence of the CH1 domain from human and mouse heavy chains

	ALA	ALA	ALA	GLU	ALA	GLY	ALA	—	GLY	GLU
	SER	LYS	SER	SER	PRO	ASP	SER	SER	SER	SER
	THR	THR	PRO	ALA	THR	LYS	THR	ILE	ALA	GLN
	LYS	THR	THR	ARG	LYS	LYS	GLN	ARG	SER	SER
	GLY	PRO	SER	ASN	ALA	GLU	SER	TRP	ALA	PHE
	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO	PRO
120	SER	SER	LYS	THR	ASP	ASP	SER	GLU	THR	ASN
	1 VAL 1	VAL	VAL	ILE	VAL	MET	VAL	LEU	LEU	VAL
	37 PHE 50	TYR	PHE	40 TYR 46	PHE	PHE	PHE	TYR	PHE	PHE
	8 PRO 17	PRO	PRO	13 PRO 15	PRO	LEU	PRO	PRO	PRO	PRO
	24 LEU 55	LEU	LEU	21 LEU 50	ILE	LEU	LEU	LEU	LEU	LEU
	3 ALA 3	ALA	SER	11 THR 11	ILE	SER	THR	LYS	VAL	VAL
	PRO	PRO	LEU	27 LEU 24	SER	GLU	ARG	PRO	SER	SER
	SER	GLY	CYS	9 PRO 15	GLY	CYS	CYS	CYS	CYS	CYS
	SER	SER	SER	PRO	CYS	LYS	CYS	—	GLU	GLU
129	5 LYS 2	ALA	THR	1 ALA 3	ARG	ALA	LYS	LYS	ASN	SER
	—	—	—	3 LEU 8	HIS	—	—	—	—	—
130	SER	ALA	GLN	SER	PRO	PRO	ASN	GLY	SER	PRO
133	THR	GLN	PRO	SER	LYS	GLU	ILE	THR	ASN	LEU
	—	—	—	—	—	—	PRO	—	—	—
134	SER	THR	ASP	—	ASP	GLU	SER	ALA	PRO	SER
	GLY	ASN	GLY	—	ASN	ASN	ASN	SER	SER	ASP
	GLY	SER	ASN	ASP	SER	GLU	ALA	MET	SER	LYS
137	THR	MET	VAL	PRO	PRO	LYS	THR	THR	THR	ASN
	—	—	—	—	—	—	—	—	—	LEU
138	ALA	VAL	VAL	VAL	VAL	ILE	SER	—	VAL	VAL
139	8 ALA 10	THR	—	15 ILE 25	VAL	ASN	VAL	—	ALA	ALA
	—	—	—	—	—	—	THR	—	—	—
140	5 LEU 8	LEU	ILE	3 ILE 9	LEU	LEU	LEU	LEU	VAL	MET
	1 GLY 3	GLY	ALA	GLY	ALA	GLY	GLY	GLY	GLY	GLY
	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
	9 LEU 37	LEU	LEU	2 LEU 29	LEU	LEU	LEU	LEU	LEU	LEU
	VAL	VAL	VAL	ILE	ILE	VAL	ALA	VAL	ALA	ALA
	12 LYS 17	LYS	GLY	1 HIS 3	THR	ILE	THR	LYS	GLN	ARG
	ASP	GLY	GLN	ASP	GLY	GLY	GLY	ASP	ASP	ASP
	TYR	TYR	PHE	TYR	TYR	—	TYR	TYR	PHE	PHE
	PHE	PHE	PHE	PHE	HIS	—	PHE	PHE	LEU	LEU
	PRO	PRO	PRO	PRO	PRO	SER	PRO	PRO	PRO	PRO
150	GLU	GLU	GLN	SER	THR	GLN	GLU	ASN	ASP	SER
	—	—	GLN	GLY	—	—	—	—	—	—
151	PRO	PRO	PRO	THR	SER	PRO	PRO	PRO	SER	THR
	VAL	VAL	LEU	MET	VAL	LEU	VAL	VAL	ILE	ILE
	THR	THR	SER	ASN	THR	LYS	MET	THR	THR	SER
154	VAL	VAL	VAL	VAL	VAL	ILE	VAL	VAL	PHE	PHE
156	SER	THR	THR	THR	THR	SER	THR	THR	SER	THR
157	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP	TRP
162	ASN	ASN	SER	GLY	TYR	GLU	ASP	TYR	LYS	ASN
	—	—	—	—	—	—	—	—	TYR	TYR
	—	—	—	—	—	—	—	—	—	GLN
163	SER	SER	GLU	LYS	MET	PRO	THR	SER	LYS	ASN
	GLY	GLY	SER	SER	GLY	LYS	GLY	ASP	ASP	ASN
165	ALA	SER	GLY	GLY	THR	LYS	SER	SER	ASN	THR
	—	—	—	—	—	—	—	—	—	GLU
166	LEU	LEU	GLN	LYS	GLN	SER	LEU	LEU	SER	VAL
167	THR	SER	GLY	ASP	SER	SER	ASN	ASN	ASP	ILE
	—	—	—	—	—	—	—	—	ILE	—
168	SER	SER	VAL	ILE	GLN	ILE	GLY	MET	SER	GLN
169	GLY	GLY	THR	THR	PRO	VAL	THR	SER	SER	GLY
171	VAL	VAL	ALA	THR	GLN	GLU	THR	THR	THR	ILE
	11 HIS 16	HIS	ARG	5 VAL 12	ARG	HIS	MET	VAL	ARG	ARG
	THR	THR	ASN	1 ASN 2	THR	VAL	THR	ASN	GLY	THR
	42 PHE 70	PHE	PHE	45 PHE 72	PHE	PHE	LEU	PHE	PHE	PHE
	5 PRO 32	PRO	PRO	10 PRO 27	PRO	PRO	PRO	PRO	PRO	PRO
	2 ALA 2	ALA	PRO	PRO 2	GLU	SER	ALA	ALA	SER	THR
177	16 VAL 44	VAL	SER	6 ALA 28	ILE	GLU	THR	—	VAL	LEU
	—	—	—	—	—	—	THR	—	—	—
178	1 LEU 1	LEU	GLN	LEU 1	GLN	MET	LEU	LEU	LEU	ARG
	12 GLN 13	GLN	ASN	5 ALA 4	ARG	ARG	THR	GLY	ARG	THR
180	6 SER 4	SER	ALA	SER	ARG	ASN	LEU	SER	GLY	GLY
182	SER	ASP	SER	GLY	ASP	GLY	SER	GLU	GLY	GLY
183	GLY	—	GLY	GLY	SER	ASN	GLY	—	LYS	LYS
	—	—	ASN	—	—	—	—	—	—	—
184	LEU	LEU	LEU	ARG	TYR	—	HIS	LEU	—	TYR
	TYR	TYR	TYR	TYR	TYR	TYR	TYR	LYS	TYR	LEU
	5 SER 14	THR	THR	2 THR 23	MET	THR	ALA	VAL	ALA	ALA
	5 LEU 10	LEU	THR	MET 10	THR	MET	THR	THR	ALA	THR
	13 SER 33	SER	SER	5 SER 32	SER	VAL	ILE	THR	THR	SER
	SER	SER	SER	ASN	SER	LEU	SER	SER	SER	GLN
190	6 VAL 21	SER	GLN	30 GLN 49	GLN	GLN	LEU	GLU	GLN	VAL

Table 1. (Continued)

	VAL	VAL	LEU	LEU	LEU	VAL	LEU	VAL	VAL	LEU
	THR	THR	THR	THR	SER	THR	THR	THR	LEU	LEU
	VAL	VAL	LEU	LEU	THR	VAL	VAL	—	LEU	SER
	PRO	PRO	PRO	PRO	PRO	LEU	—	SER	PRO	PRO
	SER	SER	ALA	ALA	LEU	ALA	SER	TRP	SER	LYS
	SER	SER	THR	VAL	GLN	SER	GLY	GLY	LYS	SER
197	SER	PRO	GLN	GLU	GLN	GLU	ALA	LYS	ASP	ILE
	—	—	CYS	—	—	—	—	—	VAL	—
198	LEU	ARG	LEU	CYS	TRP	LEU	TRP	SER	MET	LEU
	—	—	—	—	—	—	—	—	GLN	GLU
199	GLY	PRO	ALA	PRO	ARG	ASN	ALA	ALA	GLY	GLY
200	THR	SER	GLY	GLU	GLN	LEU	LYS	LYS	THR	SER
202	—	—	—	GLY	—	—	—	—	—	—
203	GLN	GLU	LYS	GLU	GLY	ASN	GLN	ASN	ASN	ASP
205	THR	THR	SER	SER	GLU	HIS	MET	GLY	GLU	GLU
206	TYR	VAL	VAL	VAL	TYR	—	PHE	—	HIS	TYR
	—	—	—	—	—	—	—	—	VAL	LEU
207	ILE	THR	THR	LYS	LYS	THR	THR	THR	VAL	VAL
	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS	CYS
	ASN	ASN	HIS	SER	VAL	THR	ARG	HIS	LYS	LYS
210	VAL	VAL	VAL	VAL	VAL	ILE	VAL	VAL	VAL	ILE
	ASN	ALA	LYS	GLN	GLN	ASN	ALA	THR	GLN	HIS
	HIS	HIS	HIS	HIS	HIS	LYS	HIS	HIS	HIS	TYR
	LYS	PRO	TYR	ASP	THR	PRO	THR	PRO	PRO	GLY
	PRO	ALA	THR	—	ALA	LYS	PRO	PRO	ASN	GLY
	SER	SER	—	SER	SER	ARG	SER	SER	GLY	LYS
	ASN	SER	ASN	ASN	LYS	LYS	SER	PHE	ASN	ASN
	THR	THR	PRO	PRO	SER	GLU	THR	ASN	LYS	ARG
218	LYS	LYS	SER	VAL	LYS	LYS	ASP	GLU	GLU	ASP
	—	—	—	—	—	—	TRP	—	—	—
219	VAL	VAL	GLN	GLN	LYS	PRO	VAL	—	LYS	LEU
220	ASP	ASP	ASP	GLU	GLU	PHE	ASP	SER	ASP	HIS
	7 LYS 12	LYS	VAL	LEU	ILE	LYS	ASN	ARG	VAL	VAL
	ARG	LYS	THR	ASP	PHE	PHE	LYS	THR	PRO	PRO
223	VAL	ILE	VAL	VAL	—	PRO	THR	ILE	LEU	ILE
226	GLU	—	—	ASN	—	—	PHE	LEU	PRO	PRO
	PRO	—	—	CYS	—	—	SER	—	—	—
	10 LYS 12	—	—	—	—	—	—	—	—	—
	5 SER 3	—	—	—	—	—	—	—	—	—
230	15 CYS 20	—	—	—	—	—	—	—	—	—

The sequences are (in order) from human IgG1 [Sequence No. 1 in Kabat *et al.* (1983)], mouse IgG1 (No. 45), human IgA (No. 34), mouse IgA (No. 60), human IgD (No. 42), mouse IgD (No. 73), human IgE (No. 31), mouse IgE (No. 70), human IgM (No. 38) and mouse IgM (No. 65). The numbering scheme (in the leftmost column) is that of Kabat *et al.* (1983). The total number of contacts in the CH1:CL pair of KOL are listed on the left and the residue surface areas buried as a consequence of the formation of the KOL CH1:CL dimer are listed on the right of some residues in the human IgG1 sequence; alongside the mouse IgA sequence are the contacts found in the McPC603 CH1:CL pair and the residue surface areas buried in the McPC603 interface. The atomic contacts given here represent the totals for each residue and were obtained from Tables 3 and 6. The surface areas are in units of square Å^2.

above and the resulting structural alignment was used to align the amino acid sequences. In the regions where insertions and deletions occur, the sequences were aligned visually to maximize homology. The other CL and CH1 sequences were aligned with those of McPC603 and KOL using a version of the program written by M. Murata (Murata *et al.*, 1985) that had been modified to use the log odds matrix values of Dayhoff *et al.* (1978) as weights in the amino acid comparisons. The alignment was then adjusted to ensure that differences in length occurred in the loop regions of the domain bilayer structure (see Tables 1 and 2).

Surface calculations

The solvent accessibility of the various structures was assessed using program MS of Connolly (1983). The solvent accessibility of individual residues was computed using program ATMSRF of S. Sheriff (Sheriff *et al.*, 1985). The van der Waals atomic radii used here were those compiled by Case and Karplus (1979); a radius of 1.5 Å was assumed for the solvent probe (water).

Computation of atomic contacts

The interactions between CL and CH1 residues were computed using program CONTAX (E. A. Padlan, unpublished). Here, a pair of atoms are designated as being in contact if they are within 1.0 Å of the sum of their van der Waals radii. The atomic van der Waals radii used were those compiled by Case and Karplus (1979).

RESULTS AND DISCUSSION

Tables 1 and 2 show the listing of the interface residues aligned with corresponding residues from other classes of antibody. Included in Tables 1 and 2 are the total number of atomic interactions involving each residue and the residue surface area buried in the CH1:CL interface of the KOL and McPC603 proteins. The corresponding quantities for NEW and J539 are very similar. By and large, the number of atomic contacts that a residue makes is paralleled by the amount of surface area that is buried as a consequence of the formation of the CH1:CL dimer.

Table 2. Amino acid sequence of the light chain constant domains of human and mouse lambda and kappa chains

	GLN	GLN	ARG	ARG		7 THR 23	THR	SER	7 SER 32
	PRO	PRO	THR	ALA		THR	GLN	VAL	4 TRP 8
110	LYS	LYS	VAL	ASP		PRO	PRO	THR	7 THR 22
	ALA	SER	ALA	ALA		2 SER 6	SER	GLU	ASP
	ALA	SER	ALA	ALA		LYS	LYS	GLN	GLN
	PRO	PRO	PRO	PRO		6 GLN 10	GLN	ASP	ASP
	SER	SER	SER	THR		SER	SER	SER	SER
	VAL	VAL	VAL	VAL		—	—	LYS	LYS
	1 THR 6	THR	PHE	1 SER 11	170	ASN	ASN	ASP	ASP
	LEU	LEU	ILE	4 ILE 16		ASN	ASN	SER	SER
	42 PHE 66	PHE	PHE	60 PHE 81		LYS	LYS	THR	THR
	PRO 4	PRO	PRO	6 PRO 24		TYR	TYR	TYR	TYR
120	PRO 6	PRO	PRO	PRO		5 ALA 29	MET	SER	7 SER 18
	9 SER 26	SER	SER	7 SER 12		6 ALA 14	ALA	LEU	10 MET 15
	2 SER 8	SER	ASP	SER		4 SER 18	SER	SER	19 SER 30
	22 GLU 26	GLU	GLU	13 GLU 14		SER	SER	SER	SER
	25 GLU 47	GLU	GLN	31 GLN 32		27 TYR 53	TYR	THR	THR 24
	LEU	LEU	LEU	LEU		LEU	LEU	LEU	LEU
	GLN	GLU	LYS	THR	180	SER	THR	THR	THR
	ALA 2	THR	SER	2 SER 3		LEU	LEU	LEU	LEU
	ASN	ASN	GLY	GLY		THR	THR	SER	THR
	4 LYS 3	LYS	THR	GLY		PRO	ALA	LYS	LYS
130	ALA	ALA	ALA	ALA		GLU	ARG	ALA	ASP
	6 THR 25	THR	SER	3 SER 15		GLN	ALA	ASP	GLU
	LEU	LEU	VAL	VAL		TRP	TRP	TYR	TYR
	6 VAL 29	VAL	VAL	6 VAL 35		LYS	GLU	GLU	GLU
	CYS	CYS	CYS	CYS		SER	ARG	LYS	ARG
	29 LEU 47	THR	LEU	40 PHE 76		HIS	HIS	HIS	HIS
	5 ILE 10	ILE	LEU	LEU	190	LYS	SER	LYS	ASN
	7 SER 18	THR	ASN	8 ASN 19		SER	SER	VAL	SER
	ASP	ASP	ASN	ASN		TYR	TYR	TYR	TYR
	PHE	PHE	PHE	PHE		SER	SER	ALA	THR
140	TYR	TYR	TYR	TYR		CYS	CYS	CYS	CYS
	PRO	PRO	PRO	PRO		GLN	GLN	GLU	GLU
	GLY	GLY	ARG	LYS		VAL	VAL	VAL	ALA
	ALA	VAL	GLU	ASP		THR	THR	THR	THR
	VAL	VAL	ALA	ILE		HIS	HIS	HIS	HIS
	THR	THR	LYS	ASN		GLU	GLU	GLN	LYS
	VAL	VAL	VAL	VAL	200	GLY	GLY	GLY	THR
	ALA	ASP	GLN	LYS		—	—	LEU	SER
	TRP	TRP	TRP	TRP		—	—	SER	THR
	LYS	LYS	LYS	LYS		SER	HIS	SER	SER
150	ALA	VAL	VAL	ILE		THR	THR	PRO	PRO
	ASP	ASP	ASP	ASP		VAL	VAL	VAL	ILE
	SER	GLY	ASN	GLY		GLU	GLU	THR	VAL
	SER	THR	ALA	SER		LYS	LYS	LYS	LYS
	PRO	PRO	LEU	GLU		5 THR	SER	SER	SER
	VAL	VAL	GLN	ARG		VAL	LEU	PHE	PHE 3
	LYS	THR	SER	GLN	210	ALA	SER	ASN	ASN
	ALA	GLN	GLY	ASN		PRO	ARG	ARG	8 ARG 16
	GLY	GLY	ASN	GLY		THR	ALA	GLY	ASN
	VAL	MET	SER	VAL		12 GLU 21	ASP	GLU	GLU
160	24 GLU 33	GLU	GLN	12 LEU 42		16 CYS 20	CYS	CYS	CYS
	THR 5	THR	GLU	ASN 7		SER	SER	—	—

The sequences are (in order) from human lambda [Sequence No. 16 in Kabat *et al.* (1983)], mouse lambda (No. 31), human kappa (No. 1) and mouse kappa (No. 23). The numbering scheme is that of Kabat *et al.* (1983). The total number of atomic contacts in the CH1:CL pair of KOL are listed on the left and the residue surface areas buried as a consequence of the formation of the KOL CH1:CL dimer are listed on the right of some residues in the human lambda sequence; alongside the mouse kappa sequence are the contacts found in the McPC603 pair and the residue surface areas buried in the McPC603 CH1:Cl interface. The atomic contacts given here are the totals for each residue and were obtained from Tables 3 and 6. The surface areas are in units of Å^2.

Details of the CH1:CL contacts are presented in matrix form in Tables 3–6. In these matrices, each element represents the number of pair interactions defined as non-bonded interatomic distances of less than a certain length between residues across the interface.

In Figs 1 and 2 we see the Fab constant domains from McPC603 and from KOL. The corresponding structures from J539 and NEW are very similar. The domain structure can be regarded as a sandwich or flattened cylinder with four strands on one surface and three on the other. The interdomain interface is formed by the interaction of the two four-stranded surfaces. In this interface there are a number of amino acid residues that make contact with residues of the opposite domain (Tables 3–6). It can be observed that there are a few residues on each chain that make most of the interchain contacts. In the case of McPC603 these include Tyr122H, Pro123H, Leu124H, Phe174H and Pro175H, as well as Phe118L, Ser121L, Glu123L, Gln124L, Phe135L, Leu160L and Ser174L. For the most part these matrices are very similar for antibodies of the same class, i.e. McPC603 resembles J539 (alpha and kappa, see Tables 3 and 4), and NEW resembles KOL (gamma and lambda, see Tables 5 and 6).

Table 3. Contacts between the constant domains of the light and heavy chains of McPC603

	Y 122	P 123	L 124	T 125	L 126	P 127	A 129	L 130	I 139	I 140	L 146	H 145	V 172	N 173	F 174	P 175	A 177	A 179	T 186	S 188	Q 190
S 116									1												
I 117					3			1													
F 118			12	10	23				12	3											
P 119				1	1	2		2													
S 121	2	5*																			
E 123	5	8																			
Q 124	31																				
S 127	2*																				
S 131											2	1									
V 133			6																		
F 135			3						2						9					3	23
N 137													1								7*
L 160																	5	5	2		
S 162															2	4*	1				
W 163																4					
T 164													2	1	2	2					
S 174													2		5						
M 175															10						
S 176															17					2*	
R 213						7	1														

In this matrix of contacts, the element c(i,j) represents the number of interacting atom pairs, one atom from residue i and the other from residue j. Atoms are designated as being in contact if the distance between them is within 1.0 Å of the sum of their van der Waals radii. The atomic van der Waals radii compiled by Case and Karplus (1979) were used in the computation of these contracts. The one-letter amino acid code (Dayhoff *et al.*, 1978) is used. The heavy chain residues are across the page and the light chain residues are down the page. The residue numbers correspond to those of Kabat *et al.* (1983). An asterisk (*) indicates that the contact involves at least one possible hydrogen bond.

For the heavy chain (Table 1), it can be seen that certain highly contacting residues are also invariant or highly conserved. These include Phe122, Pro123 and Leu124 of the first segment of the heavy chain together with Leu143, Phe174 and Pro175 in the second and third segments. In the light chain (Table 2), there are several conserved residues in the first segment including, in particular, Phe118, Glu123 and Glu124. Other conserved interface residues include Thr131, Val135 and Thr162.

While the above residues present a constant pattern that might be expected for the interaction of CH1 and CL, the remaining interface residues are quite variable, and are presumably *ad hoc* contributors to the specificity of the particular combination of CH1 and CL. Nevertheless, the area that each domain contributes to the interface which is excluded from solvent is roughly constant at about 500 Å^2 (526 for KOL, 524 for NEW, 518 for McPC603 and 607 for J539). This may be compared to the solvent excluded area of about 700 Å^2 created upon the interaction of trypsin with trypsin inhibitors (Janin and Chothia, 1976).

THE INTERFACE CAVITY

A cavity has been observed in the interface between CH1 and CL. In the case of McPC603 this is lined by

Table 4. Contacts between the constant domains of the light and heavy chains of J539

	Y 122	P 123	L 124	T 125	L 126	P 127	P 128	I 139	L 143	H 145	V 172	N 173	F 174	P 175	A 177	L 178	A 179	T 186	S 188	Q 190
S 116						1		4												
I 117						9														
F 118			14	10		29	6	8												
P 119			1	1		2														
S 121	3	2																		
E 123	7	9																		
Q 124	27									5*										
S 127	2																			
S 131									2											
V 133			7																1	
F 135			2										8						4	17*
N 137																				6
L 160														6	5	4	2			
N 161														3						
S 162													4	9*	2					
W 163														5						
T 164											3	1	9	3						
D 165														1						
S 174											3		11							
M 175													10							
S 176													9							
T 180										2										
N 212					1															
E 213					16															

(See footnote to Table 3.)

Table 5. Contacts between the constant domains of the light and heavy chains of NEW

	F 122	P 123	L 124	A 125	K 129	G 136	T 137	A 139	L 140	G 141	L 143	K 145	D 146	F 174	P 175	A 176	V 177	Q 179	S 180	S 186	L 187	S 188	V 190	K 228	S 229	C 230
S 114						1	7*																			
V 115							4																			
T 116							15																			
F 118			23	1				6	4	4													8			
P 119			1	2																						
S 121	6	6*																								
S 122																								1		
E 123	4	2																								
E 124	25										1	6														
K 129												7	3+													
T 131											2	5														
V 133			1								2											3				
L 135														28								2	3			
I 136														3												
S 137														4												
E 160																	3	7	6							
T 162																3	7									
S 165															2											
A 174														2	1											
A 175														5												
S 176														1												
Y 178											1						7			1	3	9*				
K 207						1																				
E 213					9*																				2	
C 214																								3	9	10

(See footnote to Table 2.) A plus sign (+) indicates that the contact involves a favorable electrostatic interaction.

Table 6. Contacts between the constant domains of the light and heavy chains of KOL

	V 121	F 122	P 123	L 124	A 125	K 129	A 139	L 140	G 141	L 143	K 145	H 172	F 174	P 175	A 176	V 177	L 178	Q 179	S 180	S 186	L 187	S 188	V 190	K 221	K 228	S 229	C 230
T 116							1																				
F 118				23	3		7	5	1														3				
S 121		5	4*																								
S 122																									2*		
E 123	1*	12	2																					7+			
E 124		20									5+																
K 129											4																
T 131										3	3*																
V 133				1						2												3					
L 135													22									4	3				
I 136													5														
S 137												3	4														
E 160																5	1	12	6*								
T 162														1	2	4											
S 165														2													
Q 167												6*															
A 174												2	2	1													
A 175													6														
S 176													3	1													
Y 178										4						7				5	5	6*					
T 208						5*																					
E 213																									8		4
C 214																										5	11

(See footnote to Table 5.)

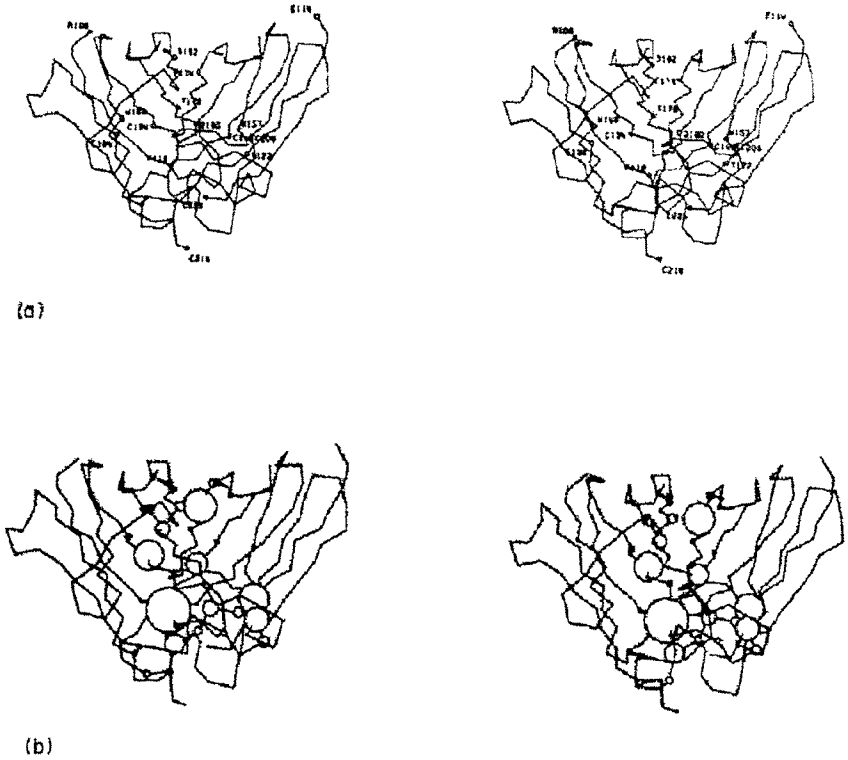

Fig. 1. (a) Stereograph of the alpha carbon skeleton of the CH1:CL domains of McPC603. The CL domain is on the left. Some residues in each chain are indicated by circles and labelled to serve as reference points. (b) The same model with circles to indicate the interacting residues. The radius of each circle is proportional to the extent of the interaction as measured by the number of pair interactions that occur between atoms of this residue and atoms of the opposite chain.

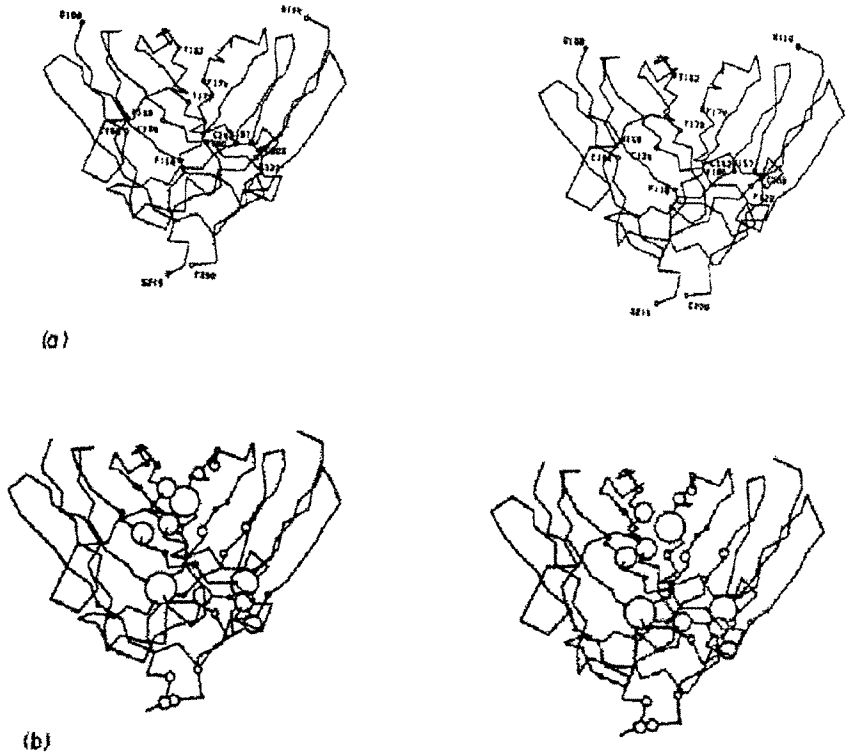

Fig. 2(a) and (b). The same as Fig. 1 but for KOL.

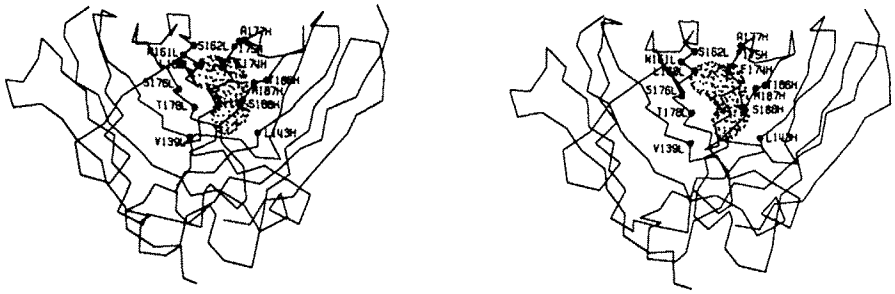

Fig. 3. Stereograph of the alpha carbon skeleton of the CH1:CL domains of McPC603 showing the location of the cavity (dotted surface) in the interface. The filled circles indicate the residues lining the cavity. The orientation is the same as in Fig. 1.

residues V133, L160, N161, S162, S176 and T178 of the light chain and L143, F174, P175, A177, T186, M187 and S188 of the heavy chain (Fig. 3). The vol of this cavity is 143 Å^3, sufficient to accommodate an aromatic side chain. Similar cavities occur in J539, NEW and KOL, with vols of 146, 70 and 49 Å^3, respectively. The cavity is presumably filled with solvent, although we have not observed any solvent density within the cavity in either McPC603 or J539, probably because of the low resolution of the X-ray data used in these analyses. It is not clear whether this cavity plays a functional role, but a possible role is suggested by the variation in the interface residues in the different isotypes, namely that it serves to provide more flexibility in the interaction between these residues. Thus any strain caused by the introduction of too bulky a side chain into the interface can be relieved by movement of side chains into the cavity.

This hypothesis was tested by modelling a composite CH1:CL structure in which the CL was from the human KOL lambda chain and the CH1 was from the mouse McPC603 alpha chain; the KOL CL being first maximally aligned with the McPC603 CL in the latter's original position relative to CH1 (see Materials and Methods). Calculation of the atomic contact between the domains of this composite structure revealed close contacts between the sidegroup of Tyr178 of CL and Ser188 of CH1. Position 178 is occupied by either Tyr or Phe in lambda chains and by Thr in kappa chains (Kabat *et al.*, 1983); the variation at this position represents the most drastic and consistent difference between homologous lambda and kappa interface residues in terms of size (Table 2). Residue 178 lines the interface cavity in McPC603 and is therefore placed with plenty of room for movement. By turning the sidegroup of Tyr178 by a mere 37° about the CA–CB bond, all the close contacts involving this residue were relieved. This reorientation positions the sidegroup of Tyr178 in the interface cavity, effectively filling most of the cavity. The larger cavity observed in the kappa:alpha pairs vs the lambda:gamma pairs can therefore be accounted for by the necessity for space in order to accommodate the threonine to tyrosine change that would occur in a kappa to lambda substitution. The presence of this cavity then permits this substitution to occur without alteration of the mode of association of CH1 and CL, this in turn permitting VH and VL to adopt their canonical quarternary structure.

Subsequently, we became aware of a similar circumstance that has been observed for T4 bacteriophage lysozyme (Alber *et al.*, 1985). There a mutation that changes Ala146 (their numbering) to threonine propagates a movement of a tryptophan residue that in turn causes the movement of the side chain of Met106, which lies in the vicinity of a cavity, causing it to move into the cavity. However, whereas this result in T4 lysozyme is coincidental, the cavity observed in the CH1:CL interface satisfies an evolutionary requirement of immunoglobulin assembly.

It would appear that the existence of internal cavities may be a natural mechanism for accommodating mutations which otherwise could cause the disruption of the structural integrity of protein molecules.

REFERENCES

Alber T., Grutter M., Gray T. M., Wozniak J. A., Weaver L. H., Baker E. N. and Matthews B. W. (1985) *UCLA Symposia on Molecular and Cellular Biology: Protein Structure, Folding and Design.* Alan R. Liss, Inc. (In press).

Bernstein F. C., Koetzle T. F., Williams G. J. B., Meyer E. F., Brice M. D., Rogers J. R., Kennard O., Shimanouchi T. and Tasumi M. (1977) *J. molec. Biol.* **112,** 535–542.

Case D. A. and Karplus M. (1979) *J. molec. Biol.* **132,** 343–368.

Connolly M. L. (1983) *J. appl. Cryst.* **16,** 548–558.

Dayhoff M. O., Schwartz R. M. and Orcutt B. C. (1978) In *Atlas of Protein Sequence and Structure* (Edited by Dayhoff M. O.), Vol. 5, Supplement 3, pp. 345–352. National Biomedical Research Foundation, Washington D.C.

Janin J. and Chothia C. (1976) *J. molec. Biol.* **100,** 197–211.

Kabat E. A., Wu T. T., Bilofsky H., Reid-Miller M. and Perry H. (1983) *Sequences of Proteins of Immunological Interest*, pp. 167–177. Department of Health and Human Services, Public Health Service, NIH, Washington DC.

Klein M., Kortan C., Kells D. I. C. and Dorrington K. J. (1979) *Biochemistry* **18,** 1473–1481.

Marquart M., Deisenhofer J., Huber R. and Palm W. (1980) *J. molec. Biol.* **141,** 369–391.

Murata M., Richardson J. S. and Sussman J. L. (1985) *Proc. natn. Acad. Sci. U.S.A.* **82,** 3073–3077.

Satow Y., Cohen G. H., Padlan E. A. and Davies D. R. (1986) *J. molec. Biol.* (Submitted).

Saul F., Amzel L. M. and Poljak R. J. (1978) *J. biol. Chem.* **253,** 585–597.

Sheriff S., Hendrickson W. A., Stenkamp R. E., Sieker L. C. and Jensen L. H. (1985) *Proc. natn. Acad. Sci. U.S.A.* **82,** 1104–1107.